Executive summary

Within the European Water Framework Directive (WFD) legislation, “good surface water chemical status” is defined as a status in which concentrations of pollutants do not exceed predefined environmental quality standards (EQS). For this purpose, a set of 55 priority substances has been defined for surveillance monitoring purposes. A chemical water quality assessment based on compound-by-compound comparisons between the concentration of an individual pollutant and its EQS ignores (1) the contribution by non-analyzed and often unknown substances to the toxic potency present in the water, (2) the combined effects of the different chemicals present in the mixture. In addition, the time-point grab sampling method as currently applied within the WFD surveillance monitoring may miss incidental higher concentrations due to peak emissions.

The TIPTOP study described in this report explored an alternative strategy for ecological risk assessment purposes, consisting of a combination of time-integrative passive sampling followed by toxicity profiling. Passive sampling is a sampling technique based on free flow of a substance from the water to a receiving sampler due to differences between chemical potentials of the substance in water and the sampler. Toxicity profiling is the testing of the sampler extract for its activity towards a battery of biological endpoints resulting in a toxicological ‘fingerprint’ of the complex mixture of substances present in the sampler. The goal of TIPTOP was to test the following two hypotheses:

H1: Compound-by-compound chemical water quality assessment based on concentrations of target-analyzed is toxicologically and ecologically less relevant and thereby less protective than risk assessment based on bioassay responses to the complex mixture of pollutants as a whole;

H2: A combination of time-integrative passive sampling followed by toxicity profiling is a more cost-effective, alternative to determine if “good surface water chemical status” is reached than the currently applied chemical analysis of a continuously expanding suite of individual substances.

Passive sampling was performed using two types of passive samplers, i.e. silicone rubber sheets as partition-based passive samplers and Speedisks as adsorption-based passive samplers. Samplers were deployed at six river sites in the Dutch delta, which were well-characterized as WFD surveillance monitoring locations, and at two urban wastewater treatment plants (WWTPs). Extracts from passive samplers were tested in a battery consisting of seven *in vitro* bioassays representing different mechanisms of action and six small volume *in vivo* bioassays with species representing different trophic levels. Passive sampler extracts were also chemically analyzed for a selected set of target compounds and for the total molar sum of accumulated compounds. Further information on compound concentrations in the water as well as ecological information about species abundance at the sampling sites was retrieved from publicly available databases and reports.

Based on the information collected, the TIPTOP results, and a combination of both, eight different strategies have been worked out to interpret the data in terms of risk assessment:

1. WFD-like comparison between concentrations in the water and environmental quality standards;
2. Calculation of toxic pressure based on concentrations in the original water phase;
3. Calculation of toxic pressure based on concentrations in passive sampler extracts;
4. Calculation of toxic pressure and HC5 based on the *in vivo* toxicity profile;
5. Calculation of minimum toxic pressure based on molar concentrations in the passive sampler extracts;
6. Benchmarking of *in vitro* and *in vivo* toxicity profiles to WWTP effluents;
7. Comparison of *in vitro* toxic potencies to mechanism-specific trigger values;
8. Calculation of mechanism-specific toxic pressure based on the *in vitro* toxicity profile.

Toxic pressure has been defined here as the probability that a concentration in the field exceeds the critical effect concentration of a species as determined in a laboratory (e.g. acute EC50, chronic NOEC). As such, toxic pressure is the same as the fraction of species potentially affected by multiple substances (msPAF), which is derived using species sensitivity distributions (SSDs).

All different risk assessment strategies led to the same conclusion that the ecological risk at the selected sampling sites is low. In approximately 3% of the cases, any of the analyzed concentrations exceeded their annual average EQS (AA-EQS). In all cases of exceedance, AA-EQS values were considered as unrealistically low, due to high safety factors (due to few toxicity data) and high bioaccumulation factors. Toxic pressure calculations based either on targeted chemical analyses in water and passive samplers or on total molar concentrations in the passive samplers also pointed out that acute EC50 values were exceeded for at maximum 0.3% of the species, regardless of the sampling location. This was confirmed by the results of the *in vivo* bioassays performed on concentrated passive sampler extracts. Based on the species-sensitivity distribution of measured acute EC50 values, the toxic pressure in non-concentrated water was assessed to be 0.0%. Using the same SSD, the concentration factor of the water was derived for which 5% of the species is estimated to be exposed above acute EC50 value. This hazardous concentration factor (HCf5), which can be regarded as a margin-of-exposure towards acute effect concentrations, ranged from 8 to 55 for the Speedisk samplers and was >100 for the silicone rubber samplers. The SSDs of measured acute EC50 values were further extrapolated into SSDs for chronic NOECs by shifting the acute SSDs by a factor of 10 to the left. Consequently, chronic toxic pressure estimates for Speedisks ranged from 0.0 to 4.9% with the exception of one of the river site locations Keizersveer (11.9%). The corresponding estimates for the margin-of-exposure towards chronic effects ranged from 1 to 5.5, with the exception again of Keizersveer, for which the margin-of-exposure was estimated as <1 (i.e. HCf5=0.8). For silicone rubber samplers, chronic toxic pressure estimates were all 0.0% with margins-of-exposure to chronic effects ranging from 12 to 53.

Apart from the generic acute and chronic toxicity determined in the *in vivo* bioassays, also the mechanism-specific toxicity determined in the mechanism-specific *in vitro* bioassays appeared to be low. Only exposure to Speedisk extracts (and not to silicone rubber extracts) resulted in a few cases where *in vitro* bioassay responses exceeded trigger values indicative for low ecological risk. For the river sites, trigger values were exceeded at maximum by a factor of two at the river sampling site of Keizersveer, indicating that ecological risk at this site cannot be excluded. Based on the same *in vitro* bioassay results, mechanism-specific toxic pressures were estimated to range from 0.1 to 31% depending on the mechanism, sampling location and sampler type considered. Highest toxic pressure was estimated for the anti-androgenic potency observed in the silicone rubber samplers only. It should be realized, however, that the calculation of anti-androgenic toxic pressure should be considered as a worst-case scenario for reasons further explained in the report.

Benchmarking of the combined *in vivo* and *in vitro* toxicity profiles observed in river sites with WWTP toxicity profiles showed that the river sites had similar toxicity profiles, except for site Keizersveer. WWTP effluents had higher toxicity profiles than the river water. In addition, most toxicity could be attributed to relatively polar compounds preferentially collected by the Speedisk samplers rather than to lipophilic compounds preferentially collected by the silicone rubber samplers.

The low toxicity observed for the passive sampler extracts in both the *in vitro* and *in vivo* bioassays hampered the evaluation of **Hypothesis 1**. Results from both chemical analysis of a limited set of target compounds and from *in vitro* and *in vivo* bioassays gave very low responses, indicating that the selected sampling sites were actually too clean to demonstrate the added value of effect-based measurements.

Since both types of measurements, including all different interpretations in terms of ecological risk assessment strategies described above, indicated little to negligible risk from chemical substances, the ultimate conclusion can be that effect-based measurements gave consistent results, with no false positives when compared to the corresponding chemical analyses.

Although it was impossible to demonstrate the additional toxicological and ecological relevance of effect-based measurements above chemical analyses, the different risk assessment strategies clearly illustrated that effect-based measurements yield more informative conclusions in terms of ecological risk assessment. In contrast to the chemical approach, effect-based measurements allow the assessment of a mixture with unknown composition and the assessment of a margin-of-exposure towards concentrations where effects in the field may be expected. This margin of exposure may be used to prioritize sampling sites for further investigation, such as the river sampling location Keizersveer, which was indicated by many different risk assessment strategies as the site with most deviating and highest toxicity profiles compared to the other river sampling locations. Moreover, the relative small fraction of the observed toxicity explained by the target analyses indicates that the observed toxicity – albeit very small – should be attributed to other compounds than the chemically analyzed compounds.

Based on the TIPTOP experiences, a stepwise strategy was proposed to determine water quality using passive samplers:

1. Characterize concentration distribution of the total molar concentration of substances collected on passive sampling devices and deduce distribution of concentrations in the water system sampled;
2. Characterize the distribution of generic acute EC50s of the most important chemical/species pairs in the water system;
3. Optionally characterize the distribution of generic chronic NOECs of the most important chemical/species pairs in the water system;
4. Optionally characterize the distributions of specific acute EC50s and chronic NOECs of the most important chemical/species pairs in the water system;
5. Calculate the generic acute toxic pressure by combining 1 and 2;
6. Optionally calculate generic chronic toxic pressure by combining 1 and 3;
7. Optionally calculate specific acute and chronic toxic pressure by combining 1 and 4;
8. Compare the result of 5-7 to standards (to be) set for ‘maximally acceptable toxic pressure’.

The costs associated with this proposed strategy requiring measurements of total molar concentrations and a number of *in vitro* bioassays on four samples per sampling location per year is estimated to be around 40% of the costs currently associated with WFD surveillance monitoring, confirming the higher cost-effectiveness of an effect-based approach. But even if the proposed strategy is not adopted, an effect-based strategy is considered as more cost-effective as it yields more information per euro spent than chemical analyses. For instance, the estimation of a margin-of-exposure towards expected effect concentrations is more informative than results from chemical analyses of dozens of compounds reported as <LOD. Moreover, it is expected that risk assessment approaches based on chemical analyses will only expand to monitoring an indefinitely large suite of chemicals. In the hypothetical case that the full set of chemicals should be monitored, it can be argued from a theoretical point of view that the number of mechanisms of action covered by these substances should always be lower than the number of substances, implying that effect-based monitoring is more cost-effective by definition. Based on these arguments, we conclude that **Hypothesis 2** is confirmed by the TIPTOP study.

With respect to effect-based monitoring using passive samplers, the following SWOT analysis was made during the TIPTOP project. **Strengths** include the cost reduction, reduced uncertainty about missed substances and missed pollution episodes, an endpoint evaluation closer to WFD-aim of good ecological status, easier interpretation in ecologically relevant terms, more informative results, and the possible use as triage method to determine hot spots for in depth study. **Weaknesses** include the misfit with current substance oriented legislation, the difficulties to attribute observed effects to underlying causation, polluting processes and sources of pollution, the need for secluded sampling stations, the need for further research into methods for sample preparation, extraction, and determination of molar concentration, the possibility that peak exposures are missed, and the theoretical impossibility to exactly translate the observed toxicity in the passive sampler to a corresponding toxicity in the water. **Opportunities** are that analytical work will not increase with future developments in pollution diversity, design and setting of effect oriented EQS procedures are less demanding and do not need regular update, and that biological triage leaves more money available to in depth study of local hot spots. **Threats** for an effect-based monitoring strategy are that changing to effect oriented legislation may take a long time, the proof of concept requires temporary simultaneous application of old and new system (i.e. temporary double costs), water quality authorities require more ecotoxicologically trained personnel, which is less available, and that technologically oriented risk assessors do believe in the outcome of chemical analyses, but not in the outcome of biological test systems.

Finally, apart from the hypothesis testing described above, additional efforts have been made within TIPTOP addressing specific aspects of passive sampling and/or toxicity profiling.

- An additional study was performed into the time-integration aspect of the passive samplers with respect to the accumulation of individual compounds, the accumulation of total molarities of compounds, and the overall toxic potency of the complex mixture sampled. For this purpose, additional time series were sampled at two sampling locations. Time-integrative sampling capacities were clearly demonstrated for the adsorption-based Speedisk samplers. Consequently, Speedisks appear a promising option for wider application in passive sampling. Such a wider application requires investment in further studies on uptake mechanism in order to control or calibrate the sampling rate, but the current results demonstrate that such an investment seems to be worth the effort.
- An additional study was performed focusing on the identification of unknown substances, which are responsible for that part of the observed toxic potency that cannot be explained by the target-analyzed compounds. For this purpose, a high-throughput effect-directed analysis (HT-EDA) approach was followed combining miniaturized bioassays with high resolution fractionation. Few bioactive compounds were identified in this study, mainly acting as estrogenic compounds. Future implementation of other detection methods could increase the number of substances to be identified.
- A zebrafish toxicity-array was developed and applied to the TIPTOP samplers. For this study, differentially expressed mRNA profiles were determined in zebrafish embryos that had been exposed to passive sampler extracts within the regular in vivo test battery. The array allows parallel screening of up to 42 target genes, which cover key processes in different toxicity pathways. As expected xenobiotic metabolism was the most affected mode of action, while no significant changes were observed for genes involved in immunotoxicity, hypoxia, general stress and methylation. Despite the need for further development, the zebrafish toxicity-array possesses the capability to become a valuable highly sensitive tool for toxicity screening of complex environmental mixtures.